



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/703,121	10/31/2000	Kasim Selcuk Candan	073303.0116	8771

7590 10/27/2004

David A Blumenthal
Foley & Lardner
2029 Century Park East
35th Floor
Los Angeles, CA 90067-3021

EXAMINER

HALIM, SAHERA

ART UNIT	PAPER NUMBER
----------	--------------

2157

DATE MAILED: 10/27/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No. 09/703,121	Applicant(s) CANDAN ET AL.	
	Examiner Sahera Halim	Art Unit 2157	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 15 July 2004.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-15 and 17-33 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☐ Claim(s) 1-15 and 17-33 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 31 October 2003 is/are: a) ☐ accepted or b) ☒ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
 2. ☐ Certified copies of the priority documents have been received in Application No. _____.
 3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|--|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948) | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____ |

DETAILED ACTION

1. This Office Action is in response to communication received on July 15, 2004.
2. Claims 1-15, and 17-33 were elected of reexamination, thus, claims 1-15 and 17-33 have examined.

Drawings

3. This application lacks formal drawings. The informal drawings filed in this application are acceptable for examination purposes. When the application is allowed, applicant will be required to submit new formal drawings.
4. The drawings are objected to as failing to comply with 37 CFR 1.84(p)(4) because reference character "40" has been used to designate both "the mirror server in Central Region and the mirror server in the Atlantic Region" on page 9, paragraph 1 of the specification.
5. The drawings are objected to under 37 CFR 1.83(a) because they fail to show "mirror server 36" as described in the specification. Any structural detail that is essential for a proper understanding of the disclosed invention should be shown in the drawing.
MPEP § 608.02(d).

6. The drawings are objected to as failing to comply with 37 CFR 1.84(p)(4) because reference character "36" has been used to designate both "the mirror server and the requests" on page 9 of the specification.

7. The drawings are objected to as failing to comply with 37 CFR 1.84(p)(4) because reference character "48" has been used to designate both "mirror server" and "reference character" on page 10 of the specification.

Corrected drawing sheets are required in reply to the Office action to avoid abandonment of the application. Any amended replacement drawing sheet should include all of the figures appearing on the immediate prior version of the sheet, even if only one figure is being amended. The replacement sheet(s) should be labeled "Replacement Sheet" in the page header (as per 37 CFR 1.84(c)) so as not to obstruct any portion of the drawing figures. If the changes are not accepted by the examiner, the applicant will be notified and informed of any required corrective action in the next Office action. The objection to the drawings will not be held in abeyance.

Claim Rejections - 35 USC § 103

8. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the

Art Unit: 2157

invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

9. Claim 1-15 and 17 – 33 are rejected under 35 U.S.C. 103(a) as being unpatentable over U.S. Pat. No. 6,175,869 to Ahuja et al (hereinafter Ahuja).

10. Ahuja discloses in a content delivery system having m servers $S = \{S_1, \dots, S_m\}$, n customers, $C = \{C_1, \dots, C_n\}$, and g geographic locations, $G = \{G_1, \dots, G_g\}$, wherein $sdel_k$ is a server delay of server S_k , $ndel_{j,k}$ is a network delay observed by customers in geographic location G_j while retrieving content from server S_k , p_i is a priority value for customer C_i , c_i is a total load of customer C_i , $u_{i,j}$ is a fraction of requests coming to customer C_i from region G_j , $a_{i,j,k}$ is a mapping representing a fraction of requests coming to customer C_i from region G_j that have been redirected to server S_k , and s_k represents a load capacity of server S_k , a method for distributing server loads, the method comprising the steps of (see abstract, Fig. 1-3):

representing an average prioritized observed response time as

$$AORT = \frac{\sum_{i=1}^n \sum_{j=1}^g \sum_{k=1}^m a_{i,j,k} \times u_{i,j} \times c_i \times p_i \times (sdel_k + ndel_{j,k})}{\sum_{i=1}^n c_i \times p_i}; \text{ and}$$

generating a mapping that assigns requests from customers to a particular server while minimizing AORT (See col. 1, line 14 – 62, col. 2, line 16 – 63 and col. 5, line 30 – 40).

Ahuja does not disclose all the variables in claims, however, Ahuja teaches parameters such average response time, total response time for each request, total end-to-end delay, and server response time (see col. 5, lines 30 – 40). It would have been obvious for a person having ordinary skill in the art at the time of the invention to recognize that all the claimed variables are similarly applied to reach the same result. A

Art Unit: 2157

person having ordinary skill in the relevant art at the time of the invention would have been motivated to modify Ahuja in order to reduce AORT (col. 2, lines 32 – 36).

11. Regarding claim 2, Ahuja discloses the step of assigning all requests from all customers in all regions to a particular server such that for each

$$C_i \in C', G_k \in G', \sum_{k=1}^K a_{i,j,k} = 1.0. \quad (\text{See col. 1, lines 22, - 36})$$

12. Regarding claim 3, Ahuja teaches the step of assigning requests to a particular server while ensuring that the load capacity of each server not exceeds such that each

for each

$$S_k \in S', \sum_{i=1}^n \sum_{j=1}^g a_{i,j,k} \times u_{i,j} \times c_i \leq s_k. \quad (\text{See col. 4, lines 64 – col. 5, line 26})$$

13. Regarding claim 4, Ahuja discloses the step of assigning request to a particular server while balancing the load of each server to within a maximum allowed deviation from a balanced state Θ such that, for all pairs of servers, S_k and S_l

$$\frac{\sum_{i=1}^n \sum_{j=1}^g a_{i,j,k} \times u_{i,j} \times c_i}{\sum_{i=1}^n \sum_{j=1}^g a_{i,j,l} \times u_{i,j} \times c_i} \leq (1 + \Theta) \times \frac{s_l}{s_k}. \quad (\text{See col. 5, line 62 – col. 6, line 50})$$

14. In reference to claim 5, Ahuja discloses a method as recited in claim 4, wherein if the content delivery system should add one or more servers, or remove one or more customers, the load of each server will not be redistributed unless the maximum allowed deviation from a balanced state Θ is exceeded (See col. 3, lines 1 – 21).

Art Unit: 2157

15. Regarding claim 6, Ahuja teaches the step of adding one

or more customers to the content delivery system only if $AORT_{new} \leq (1 + \Phi) \times AORT_{old}$:

wherein $AORT_{old}$ and $AORT_{new}$ are old and new values of $AORT$ defined

$$\text{as } AORT_{old} = \frac{\sum_{i=1}^n \sum_{j=1}^g \sum_{k=1}^m a_{i,j,k} \times u_{i,j} \times c_i \times (sdel_k + ndel_{j,k})}{\sum_{i=1}^n c_i}$$

$$\text{and } AORT_{new} = \frac{\sum_{i=1}^n \sum_{j=1}^g \sum_{k=1}^m a'_{i,j,k} \times u_{i,j} \times c_i \times (sdel_k + ndel_{j,k})}{\sum_{i=1}^n c_i};$$

wherein $a'_{i,j,k}$ is a new mapping resulting from the addition of one or more customers; and

wherein Φ is an allowable change in $AORT$ for existing clients.

(See col. 3, lines 1 – 21). Although, Ahuja does not disclose adding of customers to the content delivery system, however having the teaching Ahuja it would have been obvious for person having ordinary skill in the art at the time of the invention to include the above limitations in order to minimize $AORT$ (See col. 3, lines 1 – 21).

16. Reference to claim 7 and 8, Ahuja teaches the step of using a linear constraint solver and a non-linear constraint solver to generate the mapping (See col. 5, lines 46 – col. 6, line 50).

17. Per claim 9, Ahuja discloses using a heuristic algorithm to generate the mapping, the heuristic algorithm to generate the mapping, the heuristic algorithm comprising assigning large $a_{i,j,k}$ values to small $u_{i,j} \times c_i \times (sdel_k + ndel_{j,k})$ values to produce a smaller overall

Art Unit: 2157

AORT value (col. 5, line 47 – col. 6, line 8).

18. Regarding claim 10, Ahuja does not teach the heuristic algorithm comprising the steps of:

generating a plurality of sorted lists by sorting C_i values in increasing order of c_i , sorting $\langle C_i, G_j \rangle$ pairs in increasing order of u_{ij} , sorting S_k values in increasing order of $sdel_k$, and sorting $\langle G_j, S_k \rangle$ pairs in increasing order of $ndel_{j,k}$;

starting with a top-most, smallest value item in each list, identifying comparable smallest-value items from the other lists to generate a plurality of $\langle C_i, G_j, S_k \rangle$ triples equivalent to the number of sorted lists;

selecting from the plurality of $\langle C_i, G_j, S_k \rangle$ triples, the $\langle C_i, G_j, S_k \rangle$ triple with the smallest $u_{ij} \times c_i \times (sdel_k + ndel_{j,k})$ value;

assigning to a server S_k of the selected $\langle C_i, G_j, S_k \rangle$ triple a remaining load from the $\langle C_i, G_j \rangle$ pair; and

repeating the heuristic algorithm starting with generating the plurality of sorted lists, taking into account the changes in the values of the C_i values and the $\langle C_i, G_j \rangle$ pairs as a result of the previous server assignment during each iteration, until the load from all $\langle C_i, G_j \rangle$ pairs has been assigned to a server S_k ;

wherein if, during any iteration of the heuristic algorithm, the load capacity of the server S_k is not sufficient to handle the remaining load, the remaining load capacity of the server S_k is assigned to some of the load of the $\langle C_i, G_j \rangle$ pair, and an unassigned portion of the load from the $\langle C_i, G_j \rangle$ pair is reinserted into the iterative process.

However, it would have been obvious for one having ordinary skill in the art at the time of the invention to use the heuristic algorithm in any form in order to balance the load of the system and minimize AORT (See col. 2, line 64 – 20).

Art Unit: 2157

19. Regarding claim 11, Ahuja does not teach the heuristic algorithm comprising the steps of:

generating a load-capacity prioritized sorted list by sorting C_i values in decreasing order of remaining load capacity s_k ;

starting with a top-most, largest value item in the load-capacity prioritized list, identifying comparable smallest-value items from the other lists to generate a load-capacity prioritized $\langle C_i, G_j, S_k \rangle$ triple;

considering the load-capacity prioritized $\langle C_i, G_j, S_k \rangle$ triple in the selection of the top-most $\langle C_i, G_j, S_k \rangle$ triple with the smallest $u_{i,j} \times c_i \times (sdel_k + ndel_{j,k})$ value; and

repeating the heuristic algorithm starting with generating the plurality of sorted lists, taking into account the changes in the values of the C_i values, the $\langle C_i, G_j \rangle$ pairs, and remaining load capacity as a result of the previous server assignment during each iteration, until the load from all $\langle C_i, G_j \rangle$ pairs has been assigned to a server S_k .

However, it would have been obvious for one having ordinary skill in the art at the time of the invention to use the heuristic algorithm in any form in order to balance the load of the system and minimize AORT (See col. 2, line 64 – 20).

20. Regarding claim 12, Ahuja does not teach the heuristic algorithm comprising the steps of:

generating a list of content-available $\langle C_i, S_k \rangle$ pairs in which the content of customer C_i is stored in server S_k ; and

selecting the $\langle C_i, G_j, S_k \rangle$ triple with the smallest $u_{i,j} \times c_i \times (sdel_k + ndel_{j,k})$ value that is also part of the list of content-available $\langle C_i, S_k \rangle$ pairs;

wherein if, during any iteration of the heuristic algorithm, there is no $\langle C_i, G_j, S_k \rangle$ triple that is also part of the list of content-available $\langle C_i, S_k \rangle$ pairs, a suitable $\langle C_i, G_j, S_k \rangle$ triple with the smallest $u_{i,j} \times c_i \times (sdel_k + ndel_{j,k})$ value is chosen and the data of customer C_i is migrated to server S_k .

However, it would have been obvious for one having ordinary skill in the art at the time of the invention to use the heuristic algorithm in any form in order to balance the load of the system and minimize AORT (See col. 2, line 64 – 20).

21. Per claim 13, Ahuja does not teach the heuristic algorithm further including the step of generating a list of content-unavailable $\langle C_i, S_k \rangle$ pairs in increasing order of migration time penalty for which the content of customer C_i is not stored in server S_k ;

wherein if, during any iteration of the heuristic algorithm, there is no $\langle C_i, G_j, S_k \rangle$ triple that is also part of the list of content-available $\langle C_i, S_k \rangle$ pairs, a suitable $\langle C_i, G_j, S_k \rangle$ triple with the smallest combined $u_{i,j} \times c_i \times (sdel_k + ndel_{j,k})$ value and $\langle C_i, S_k \rangle$ migration time penalty is chosen and the data of customer C_i is migrated to server S_k .

However, it would have been obvious for one having ordinary skill in the art at the time of the invention to use the heuristic algorithm in any form in order to balance the load of the system and minimize AORT (See col. 2, line 64 – 20).

22. Regarding claim 14, Ahuja teaches claim 1 including the steps of estimating

Art Unit: 2157

$ndel_{j,k}$ for non-persistent connections using HTTP logs, the step of estimating $ndel_{j,k}$ for non-persistent connections using HTTP logs comprising the steps of:

computing an estimated round trip delay $\Delta r_{server,client}$ as $t_{con_req_rec}(i+1) - t_{resp_send_end}(i)$ from information stored in the HTTP logs, where $t_{con_req_rec}(i+1)$ represents a time at which a connection request message is received by the server for an $(i+1)^{th}$ object, and $t_{resp_send_end}(i)$ represents a time at which the server stops sending an i^{th} object; and
 computing the response time as

$(t_{resp_send_end} - t_{con_req_rec}) + 2 \times \frac{\Delta r_{server,client}}{2}$, where $t_{con_req_rec}$ represents a time at which a

connection request message is received by the server for an object, and $t_{resp_send_end}$ represents a time at which the server stops sending the object (See col. 4, lines 14 – 34, col. 5, lines 27 – 45, col. 6, line 51 – col. 9, line 35).

Ahuja does not disclose all the variables in the claim, however, Ahuja does disclose minimizing an average response time of each server taking into account the address information and load of each server (See col. 2 line 14 63). It would have been obvious for a person having ordinary skill in the art at the time of the invention to recognize that all the claimed variables are similarly applied to reach the same result, which is to minimize average response time. A person having ordinary skill in the relevant art at the time of the invention would have been motivated to modify Ahuja in order to improve network performance (col. 3, lines 1 – 20).

23. As to claim 15, Ahjua teaches claim 1 including the steps of estimating

Art Unit: 2157

$ndel_{j,k}$ for persistent connections using HTTP logs, the step of estimating $ndel_{j,k}$ for persistent connections using HTTP logs comprising the steps of:

computing an estimated round trip delay $\Delta r_{server,client}$ as $t_{con_close_rec} - t_{resp_send_end}(last)$, where $t_{con_close_rec}$ represents a time at which the server receives a request to close the persistent connection, and $t_{resp_send_end}(last)$ represents a time at which the server stops sending a response for a last request; and

computing the response time as

$(t_{con_close_rec} - t_{con_req_rec}) + \frac{\Delta r_{server,client}}{2} - \frac{\Delta r_{server,client}}{2}$, where $t_{con_req_rec}$ represents a connection request time.

Ahuja does not disclose all the variables in the claim, however, Ahuja does disclose minimizing an average response time of each server taking into account the address information and load of each server (See col. 2 line 14 63). It would have been obvious for a person having ordinary skill in the art at the time of the invention to recognize that all the claimed variables are similarly applied to reach the same result, which is to minimize average response time. A person having ordinary skill in the relevant art at the time of the invention would have been motivated to modify Ahuja in order to improve network performance (col. 3, lines 1 – 20).

24. Independent claims 17 and 18 have similar limitations as to claims 14 and 15, thus they are rejected under the same rational.

25. claims 19 – 33 have similar limitations as claims 1- 15, thus they are rejected under the same rational.

Conclusion

26. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

U.S. Pat. No. 6,314,465 to Paul et al.

U.S. Pat. No. 6,256,675 to Rabinovich, Michael

U.S. Pat. No. 5,924,116 to Aggarwal et al.

U.S. Pat. No. 6,298,381 to Shah et al.

U.S. Pat. No. 6,052,718 to Gifford, David K.

27. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Sahera Halim whose telephone number is (703) 305-8054. The examiner can normally be reached on M-F from 8:30-5:00.

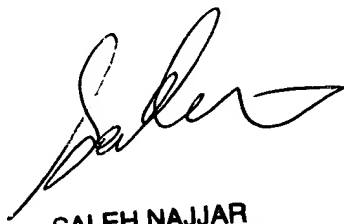
If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Ario Etienne can be reached on (703) 308-7562. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Art Unit: 2157

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

Sahera Halim
Patent Examiner
AU: 2157

October 19, 2004



SALEH NAJJAR
PRIMARY EXAMINER